CURRICULUM BASED MEASUREMENT NORMING FOR MATH (CALCULATION) IN SCHOOL DISTRICT 57

By

Gail B. Walraven

B. Ed. (Elem.) University of British Columbia, 1972

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF EDUCATION

In

CURRICULUM AND INSTRUCTION

THE COLLEGE OF ARTS, SOCIAL, AND HEALTH SCIENCES

THE UNIVERSITY OF NORTHERN BRITISH COLUMBIA

April 2001

UNIVERSITY OF NORTHERN BRITISH COLUMBIA LIBRARY Prince George, BC

TABLE OF CONTENTS	
Abstract	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
INTRODUCTION	1
Historical Background of CBM	2
CBM in School District #57	2
Need for Local Curricula and Norms	4
Results of CBM Reading and Writing Project	6
CBM Math (Calculation) Norming Project	7
Personal and School District Significance	9
Theoretical Significance	9
Need for Further Research	10
Scoring Issues	11
Scope of the Current Project	12
METHOD	13
Subjects	13
Instruments	13
Procedure	14
Controlling School Effect	15
Ethics Approval	15
RESULTS	16
Preliminary Analysis	16
Demographic Analysis	16
Problems in the Data	18
Main Analysis	19
Descriptive Statistics by Grade	19
Reliability of Measures	21
Analysis of Probe Difficulty	22
Analysis of Probe Difficulty Using ANOVA and Rank Order	23
Analysis of Probe Difficulty Using Box Plots	27
Summary of Probe Difficulty Differences	27
Creation of the Norming Tables	28
Smoothing	28

DISCUSSION	N	31
Issues	Raised by Data	31
	Students' Growth and Summer Effect	31
	Reliability of the Measures	32
~ .	Probe Difficulty	33
Concl	uding Statement	34
References		36
Appendix A	Problem Solving Approach and Process Chart	39
Appendix B	Grade 7 CBM Math Probe	41
Appendix C	Student Selection/ Probe Sequence Table	44
Appendix D	Random Selection of Students Table	46
Appendix E	Letters of Permission	48
Appendix F	Box Plots for all Norming Periods	51
Appendix G	Percentile Scores – Raw Data Grades 1-7	61
Appendix H	Charts of Smoothed Percentile Data Grades 1-6	66
Appendix I	CBM Math Norms Grades 1-6	73

iv

INTRODUCTION

Formal assessment has been important to modern educators since the end of the nineteenth century (Salvia &Ysseldyke, 1991); however, as the twenty-first century begins, it has become important to many people other than educators. Members of the general public and politicians are calling for accountability in public education and as a result schools are being more closely scrutinized than ever before. There has been an increased emphasis on assessment, as it is believed by some that testing students is the best way to determine whether or not education tax dollars are being well spent. There is no disputing the fact that assessment is important in education. According to Deno (1985), "measurement of student achievement is basic to evaluating the success of our educational programs" (p. 219). Determining how to measure student achievement is where opinions differ.

Commercial standardized tests have been popular in the United States and Canada since E. L. Thorndike's work early in the last century. By the late 1930s over 4000 tests were available for use in schools (Worthen, Borg, & White, 1993). Recently researchers and educators have begun to argue against the use of commercial standardized norm-referenced tests. In addition to the cost, educational criticisms of this kind of test include their being too broad or too biased in curricular content (Deno, 1985), failing to describe growth and a reliance on face validity (Fuchs & Fuchs, 1992), and not being useful for either diagnosing students' educational problems or for planning educational programs (Brigance & Hargis, 1993; Salvia & Ysseldyke, 1991). These inadequacies of commercially prepared tests have led school districts to create their own alternative assessment tools (Deno, 1985). Curriculum-based measurement (CBM) is an assessment tool that is tied to local curricula and can be used both to diagnose educational problems and to plan educational programs.

Historical Background of CBM

Between 1977 and 1983, Stanley Deno and others at the University of Minnesota's Institute for Research on Learning Disabilities (IRLD) conducted research for the purpose of narrowing the gap between measurement and instruction. Based on their research, they believed that to evaluate student achievement, teachers often relied more on information obtained through informal observation rather than on the results of commercial standardized tests. They found this to be problematic because of the poor reliability and validity of informal observation as an assessment tool. Curriculum-based measurement was the set of procedures that Deno (1985) and his colleagues developed as a result of their research to provide a valid and reliable alternative to teacher observation or standardized testing. It was originally designed to be used by special education teachers, but is now also used by general education teachers (Shinn & Bamonto, 1998). Knutson and Shinn (1991) define CBM as "a set of standardized, specific procedures designed to quantify student performance in basic academic skills" (p. 372). CBM procedures are reliable and valid, easy to administer, time efficient, cost effective, and can be repeated frequently (Deno, 1992). School District #57, Prince George (SD57) has adopted CBM as one assessment tool that can effectively be used by teachers and special education personnel.

CBM in School District #57

In the early 1990s, SD57 staff decided to reorganize the district's system for delivery of special education services to students. The School Support Services Task Force (SSSTF) was created for this purpose. It was in researching how other school districts were delivering services to special education students that members of the SSSTF first became interested in CBM. As part of this new delivery system, a formal four-level problem solving process was adopted. Iowa's Heartland Area Education Agency's process was used as the model for SD57. The four levels are

shown in the chart in Appendix A (School District #57, 1996b). The labeling on the chart identifies the personnel that need to be involved at each level. At all four levels of the model interventions are implemented to try and solve the student's problem before it becomes more entrenched. The student needs to be assessed frequently to determine whether the interventions are working. At the highest level of the model, Level IV, a student's eligibility and need for services beyond the resources available at the school need to be evaluated. In order to accomplish this, SD57 divided the district into fives zones and created an Area Resource Committee (ARC) in each zone. The membership of each ARC included a district assistant superintendent, an elementary school administrator and two others (usually a teacher and administrator, or two teachers). The ARC was responsible for determining whether a student who had reached Level IV was eligible for special services beyond the resources available at the school level. Determining this need "requires documentation of sufficient assessment information" (School District #57, 1996b, p. 20). In order to meet the requirement for documentation that is built into the problem-solving model, in 1993 the SSSTF recommended, "that simple informal and formal systems for gathering information for the purpose of monitoring and evaluating the effectiveness and efficiency of service delivery be developed" (School District #57, 1996a, p.3). The simple and informal system that the district adopted was CBM and it became an integral part of the assessment component of the problem solving process. The school district's decision to adopt CBM was research based. According to Salvia and Ysseldyke (1991), assessment must be purposeful. It must be more than simple observation or the collection of data. They believe that the two critical purposes of assessment are "(1) specifying and verifying problems and (2) making decisions about students" (p. 3). The decisions to be made about students may be about eligibility for, or placement in programs, or it may be

about designing effective instruction. Clearly, SD57's use of CBM data as a critical element of the problem solving process meets both components of Salvia and Ysseldyke's purposeful definition of assessment.

During the 1995-96 school year, SD57 completed a project to develop local norms for CBM in reading fluency and written expression. This norming project was conducted by a joint School District – University of Northern British Columbia (UNBC) Committee (School District #57, 1996a). Dr. Peter MacMillan carried out the research portion of the work including conducting the data analysis, establishing the norms, and writing the draft and final technical reports for this project. These norms were distributed to all elementary schools in the district and training on how to use them was provided to administrators, learning assistance teachers, and classroom teachers. The work of the joint committee culminated in the publication of a guidebook for using CBM in SD57. The norms that were established during this project were implemented more than six years after staff of SD57 had first expressed interest in CBM. Need for Local Curricula and Norms

Recall Salvia and Ysseldyke (1991) question the use of commercially prepared standardized tests as tools for diagnosing problems and planning programs. They recommend that as an alternative, "teachers and diagnosticians construct criterion-referenced achievement tests that closely parallel the curricula that the students follow" (p. 590). The term CBM includes the word curriculum because it is important to use local curricula in these measures. As Deno (1992) points out "the procedures are not curriculum based until they are applied to specific curricula" (p. 8). Several probes to be administered to students are developed at each grade level using local curricula, as this is the element that distinguishes CBM from traditional psychoeducational measurement (Deno & Fuchs, 1988). In reading, the probes are short passages taken from grade level material that students read for one minute. Students receive a score based on the total number of words read. In written expression, the probes are selections of writing produced by students after having been given the opening sentence of a story. Students write for three minutes and two scores are obtained: total number of words written, and total number of words spelled correctly. In math, probes are sets of thirty computational items that students are given five minutes to work on. Scores on the math probes are the total number of correct digits placed in the correct location. An example of a Grade 7 math probe with correct answers and scoring is included in Appendix B.

Once students' scores are obtained on a CBM probe, to what should they be compared? Applicability of CBM will depend on school districts developing local norms (Howell, Fox, & Morehead, 1993). Local norms can be developed at three levels, ranging from lowest to highest: classroom, school, and district. Kaminski and Good (1998) suggest that if possible, norms be developed at the highest level that can be managed and that norms be developed for three periods: fall, winter and spring. As the level of norming moves up from classroom to district, there is an increase in cost and effort required in developing norms. However, even at the district level, the development of local norms is a reasonable and worthwhile undertaking. The utility of the norms increases correspondingly with each level. If norms are established at the classroom level, their utility is limited. The norms can only be used for students enrolled in that one classroom. In addition, CBM data collected using classroom norms can only be used at the first three levels of the four-level problem solving process. If district norms are established, they can be used at all four levels of the problem solving process. (Shinn, Nolet, & Knutson, 1990). In order to make use of CBM as an integral part of the problem solving process, SD57 decided to establish district norms for CBM probes.

Results of CBM Reading and Writing Project

The school district's project to establish local norms for CBM in both reading fluency and written expression took place during the 1995-96 school year. In addition to the creation of the norms in these two curricular areas, the results of this project provided evidence of reliability of the measures and consistency in the difficulty of the probes. Coefficients of stability over a six-month period (October to April) for Grades 2-7 ranged from .77 to .86 in reading fluency and from .48 to .62 in written expression. Three techniques were used to analyze the probes for differences in difficulty: one factor ANOVA followed by the Scheffé post hoc test, comparison of rank order over norming periods, and an examination of box plots for lack of overlap. After all three techniques had been used, the probes were judged to be equal and norming tables by grade were created and presented to the school district (MacMillan, 1996; School District #57, 1996a).

During the 1996-1997 school year, schools in SD57 began to assess students' reading fluency and written expression using the locally developed norms. Using the <u>FileMaker Pro</u>[©] database, a SD57 teacher developed a computer file for schools to use to simplify CBM data collection and interpretation. Students' scores are entered and the program automatically converts them to the corresponding percentile score and designation in relation to average (see Appendix I). When first developed, this computer file was used to collect and interpret reading fluency and written expression scores. It was later revised to include math scores.

Following the successful implementation of the use of the CBM norms for reading fluency and written expression in 1995-96, SD57 decided to broaden its use of CBM by developing local CBM norms in another curricular area. In 1999 SD57 began a second joint project with UNBC to develop local CBM norms for computational ability in mathematics.

CBM Math (Calculation) Norming Project

7

In partnership with Dr. Peter MacMillan of UNBC, SD57 recently completed a project to create local norms for CBM in mathematics. In the summer of 1999, seven probes for each grade (1-7) were constructed using the British Columbia mathematics curriculum. For each learning outcome that represented a computational skill a set of items that tested the particular skill was constructed. For each Grade 1 probe, thirty items from the pool of items created for that grade level were randomly selected. The thirty items selected for Grades 2-7 probes included not only randomly selected items at grade level, but also several items from the previous grade. Each probe at all grade levels included items from the entire year's curriculum. Responses on the probes were scored as the number of correct digits (CD) rather than as either right or wrong. This allowed credit for partial answers.

During the 1999-2000 school year, probes were administered to a sample of 20% of Grade 1-7 students in SD57. Previous CBM norming studies have suggested that samples of 15-25% of the population are optimal (School District #57, 1996a; Shinn, 1988; Tindal, Germann, & Deno, 1983). Using the 20% sample size resulted in the selection of approximately 275 students from each grade level. This number is well above the minimum of 100 students that Shinn (1988) recommends as being "highly desirable for determining stable percentile ranks" (p. 66). Probes were administered in October, January, and April for students in Grades 2-7. As in the earlier project to develop local norms for reading fluency and written expression, Grade 1 students were not included until the April norming period. This decision was made because early in the school year these young students "have not had sufficient instruction to master basic skills, and therefore ... CBM tasks may be insensitive to student performance" (Shinn, 1989, p. 99). From the beginning, I was involved in all phases of the school district's CBM math norming project. I attended all meetings and working sessions held to create the probes and to develop the instructions and scoring rules for teachers. After SD57 collected the data I analyzed the data, established local norms, and wrote the draft and final technical reports for this project. The norms and technical reports were delivered to SD57 in August 2000.

As in the project to establish CBM norms for reading fluency and written expression, student data on the following variables were collected during each of the three norming periods: gender, age, grade, score, and probe number. In analyzing the data and developing the norms, only three variables were used: the dependent variable: score, measured as correct digits, and the independent variables: grade and probe number. First Nations status and program (regular, French Immersion, or Montessori) were variables introduced for the Math norming project. In the planning stage of the project. I suggested to SD57 personnel that data on program be collected in order to facilitate future research on CBM. It was a school district decision to collect data on First Nation status. This was done with the knowledge and consent of the Aboriginal Education Board. The aboriginal community through their board had asked that the school district collect First Nations data in order for them to have measurements of how well initiatives in aboriginal education were working. The SPSS statistical package (SPSS Graduate Pack 9.0 for Windows, 1999) was utilized to conduct all statistical analyses to determine probe difficulty and the equivalence and stability of the scores over time and across probes. Norms were also established for each grade level for each norming period. A guidebook for using CBM in math will be published by SD57 personnel.

Personal and School District Significance

My interest in becoming involved in the CBM math norming project was for personal, professional reasons. SD57 has endorsed CBM for use in elementary schools. I am employed by SD57 as an elementary zone vice principal. One of my responsibilities is to assist schools in implementing district policies and initiatives. I will be able to assist teachers and administrators in the use of CBM more effectively if I am able to increase my understanding of it. I was a member of an ARC in the district for several years. Each ARC is responsible for distributing special education funding to support students with learning and behavioral difficulties. When a school applies for funding from ARC for a particular student, CBM is one of the measures used in establishing whether the student's academic performance is sufficiently discrepant from that of his or her peers to warrant extra support at the district level. This project will be of specific interest to SD57 as it continues to implement and promote the use of CBM.

Theoretical Significance

The school district decided to establish CBM norms in reading and written expression before doing so in mathematics. At the time SD57 personnel were investigating CBM, more research had been done in reading CBM than in math. Deno (1985) believed that this was "partially because the functional purpose of reading – to obtain meaning from text - is clearer than the functional purpose of mathematics" (p. 230). Researchers at the University of Minnesota's IRLD had studied reading and writing CBM and found that they were reliable and valid assessment tools. By the late 1980s attention on CBM was still focused on reading and written expression. For example, Deno and Fuchs (1988) described in detail how to set up CBM systems and omitted information on mathematics. In Marston (1989) still felt that there were "limited math technical adequacy data subsequent to the IRLD research" (p. 51). My project will

be of interest to researchers interested in CBM across North America as it will add to the limited body of CBM knowledge in the subject of mathematics.

This project will also contribute to the body of local CBM knowledge. Three studies have already been conducted using the school district data set from the reading and writing CBM norming project (Fewster, 2000; Hedekar, 1997; MacMillan, 2000). Hedekar (1997) researched relative age and gender effects in reading fluency and written expression as measured by CBM. Hedekar found that females outperformed males at every grade level for number of words read correctly, number of words written, and number of words spelled correctly. She found that relative age was not significant for either males or females. Fewster (2000) conducted a predictive validity study. She determined that students' scores on reading and writing CBM probes in elementary school were a good indicator of students' performance in Grade 8, 9, and 10 humanities courses. MacMillan (2000) confirmed Hedekar's results on gender differences and relative age when he reanalyzed the same data set applying multi-faceted Rasch to the reading scores. The school district data set for the CBM math norming project has already been used for further local research. A UNBC Master's candidate is planning to replicate Hedekar's research using the data collected during the CBM math norming study (B. J. Foulds, personal communication, September 18, 2000), and Dr. Peter MacMillan is presenting the math results at the Canadian Society for Studies in Education Conference at Laval University in May 2001 (P. MacMillan, personal communication, January 10, 2001).

Need for Further Research

The literature discussed earlier in this paper was presented to provide an historical background and a context for CBM. It was not intended to be a comprehensive review of the literature on CBM. The many studies conducted on CBM suggest that the measures are reliable and valid for reading fluency, written expression, and spelling. Even though there have been fewer studies on CBM in math, those that have been conducted suggest that CBM math probes are reliable (Tindal, Marston, & Deno, 1983) but some types of validity may be problematic (Marston, 1989). Content validity is high as the items for the probes are taken directly from local curriculum. In a research project designed to develop local CBM norms for the Pine County Special Education Cooperative in Minnesota, Tindal, Germann and Deno (1983) found that validity coefficients were low when CBM scores were compared to two standardized math tests. This may be because the standardized tests include items that measure more than computational ability. There is a need for further research to establish reliability ratings for CBM in mathematics (Deno, 1992; Marston, 1989). Specifically Marston (1989) believes additional research is needed "to determine if the aggregation of multiple math probes significantly improves the reliability of the measures" (p. 53).

Scoring Issues

CBM math probes are scored by counting the number of correct digits located in the correct place (see Appendix B). In some of the more complex computations, raters need to make judgments as to whether or not to give credit for a digit. For example, the correct digit may be present, but in the wrong column. Moore and Young (1997) report that attaining "high rater reliability is quite possible and feasible" (p. 11). They find this to be especially true when there are specific scoring guidelines established and raters are trained. SD57 has done both these things in trying to ensure high inter-rater reliability. In 1983 three types of reliability (test-retest, alternate form, and inter-rater) were tested in spelling and math CBM. High inter-rater reliability coefficients were reported in math, ranging from .90 to .99 when the skills were broken down by type of computation (Tindal, Marston, & Deno, 1983).

Scope of the Current Project

The project to develop local CBM norms for mathematical computation was a replication of the 1995-96 project to develop norms for reading fluency and written expression. The same method for selecting subjects, using and analyzing the probes, and developing the norms was used. In analyzing the results for this project the same statistical analysis was used to address the issues of reliability of the measures and consistency in probe difficulty. It was not within the scope of this project to address the issue of inter-rater reliability. Analysis of the data collected on gender, relative age, and program is also not included in this project. However, the data on these other variables is in the data set and may be used for future research.

METHOD

The design, method, and data collection for the norming study were developed and carried out by SD57 in a replication of the 1996 reading fluency and written expression norming study. Data for this project were collected from all 52 elementary schools in School District #57, Prince George. Instructions for the selection of subjects, the selection of probes, and administrative procedures were included in an instruction manual (School District #57, 1999). <u>Subjects</u>

The sample for this study comprised approximately 20% of the students from each grade level (1-7). A stratified random sampling method was used. Twenty percent of the students were randomly selected at each grade level and within special programs. Most schools only enrolled students in the regular program. Five schools enrolled students in either the French Immersion or the Montessori program as well as in the regular program. From each school's computer management program, alphabetical lists of students were generated by grade and by grade within program. Students were selected from these lists using Student Selection/Probe Sequence table (Appendix C) in conjunction with the Random Selection of Students Table (Appendix D). The first table indicated how many students were to be selected from each school; the second table indicated which ones to select.

Instruments

The 30-item probes described earlier were the instruments used in this project. Seven were created for each grade. However, only six were assigned to schools to specifically use. The Student Selection/Probe Sequence table was used to determine which probe to use during the October norming period for Grades 2-7, and for the April norming period for Grade 1. For Grades 2-7, the next two probes were used for January and April. For example, if Probe 5 were assigned for October, then Probe 6 and 1 would be used in January and April respectively. The seventh probe that was created for each grade level was called Probe Extra. It was never assigned. It was to be given to students who completed all thirty of the items on their assigned probe before the five-minute testing period elapsed.

Procedure

Personnel from each school received training at a one-day workshop in September 1999 on how to select students, how to administer and score the probes, and how to record and submit the data. In most cases the school's Learning Assistance Teacher (LAT) was responsible for administering the probes to the selected students; however, in smaller schools a classroom teacher was responsible. A binder that included all instructions, all the probes, and all the answer keys for the probes was provided to each school (School District #57, 1999). Probes were administered to the selected students in October 1999, January 2000 and April 2000. The data collected were recorded at each school in a <u>FileMaker Pro</u>© database file that was then forwarded electronically to SD57 personnel at the school board office. In May, all hard copies of the administered probes and summary data collection sheets were also forwarded to the school board office. (Some were not returned.)

The individual <u>FileMaker Pro</u>[©] files from each school were compiled into one large file and forwarded electronically to me. I screened and cleaned the data before transferring it into the SPSS computer program for data analysis. In this screening and cleaning process each record in the file was checked for inconsistencies in the scores over the three norming periods. Hard copies of the probes were referred to when records were found to have one score that seemed to be unusually high or low when compared to the other two scores in the record. Twenty-seven records were changed during this process. In one case the total possible score had been recorded for all students in Grade 7 at one school instead of the score each child had actually received. Several copies of both the <u>FileMaker Pro</u>[©] file and the SPSS file have been made and are in different locations both at UNBC and SD57 to prevent loss of or damage to the data.

Controlling School Effect

SD57 is a large district with great variation in both size and location of its schools. There are several large urban schools with student populations of nearly 400 students. In contrast there are a few rural schools enrolling fewer than 30 students. It was possible to have a school effect if care was not taken in grouping schools to ensure that within each group urban, rural, large, and small were included. As was done in the earlier reading and writing norming project (School District #57, 1996a), this issue was dealt with by stratifying the schools and assigning probes to each school for each norming period (see Appendix C). In examining the groups of schools in Appendix C, readers with a knowledge of SD57 will notice that inner city schools and schools with special programs have been spread over the six groups.

Ethics Approval

I was involved in the CBM Math (Calculation) project in two capacities: as a SD57 employee and as a UNBC Master's degree candidate. As an employee I assisted other employees of the district to design and implement the project. I was granted permission to use the data set for my Master's degree project from the school district. I also received ethics approval from the university. Copies of these approvals are attached in Appendix E.

RESULTS

Preliminary Analysis

Demographic Analysis

The total sample consisted of 2038 students from 52 schools. There are 2039 records in the data set as one student is listed at different schools in different norming periods. Schools with students in the French Immersion or Montessori programs randomly selected students from within each program and submitted their data separated by grade and program. Records having only April data are broken down into the 282 Grade 1 students who were not included in the project in October and January, and the 35 other students in Grades 2-7. Most of the students in the sample were tested in all three norming periods. Some students were present for only one or two of the norming periods. Table 1 depicts this information.

Table 1

October	January	April	Total
V	\checkmark	V	1557
V			50
\checkmark	1		48
\checkmark		1	13
	~		5
	V	1	49
		Grade 1	282
		Grades 2-7	35
Total			2039

Students Present at Norming Periods

All records submitted were complete for data on gender. As was the case in the reading fluency and written expression project, the group of subjects was split almost equally by gender. In the earlier project the percentages were 50.9 male and 48.9 female with .2 percent missing gender data (MacMillan, 1995). In this math norming study the percentages for male and female students were 51.1 and 48.9 percent respectively.

Table 2 shows that schools accurately selected the correct number of students as specified on the Student Selection/ Probe Sequence table (Appendix C). The percentages by grade for each norming period are within between .01 and .27 of the means. One Grade 2 student's record was removed from the data file, as his scores were problematic. He had a very high score in the October norming period and very low scores in January and April. However, his school was one of the twelve that did not return some or all of the hard copies of the probes, so verification was impossible.

Table 2

Grade	October	Percent	January	Percent	April	Percent
1					282	14.55
2	278	16.66	278	16.76	275	14.19
3	280	16.79	277	16.69	279	14.40
4	278	16.66	276	16.64	274	14.14
5	281	16.85	276	16.64	277	14.29
6	276	16.55	277	16.69	276	14.24
7	275	16.49	275	16.58	275	14.19
Total	1668	100.00	1659	100.00	1938	100.00
Mean		16.67		16.67		14.28

Number of Students by Grade

The Student Selection/ Probe Sequence table (Appendix C) also indicated which probe each school was to use for the October norming period. Schools were to then use the next two probes in sequence in the January and April norming periods. The percentages in Table 3 are not as close to the means as the percentages in Table 2 were. The percentages in Table 3 are within .06 and 3.04 of the means. There are two possible explanations for this. First, in totaling the number of students to be selected from each grade by groups of schools as identified in Appendix C, there is a range of between 43 and 50. (In totaling the first group of schools, I used the average of 3 for Seymour's Montessori program.) Second, when I referred to the hard copies of the probes, I discovered that some schools did not use the probes as assigned. The data in Table 3 shows that the group of schools that administered Probe 3 in October, Probe 4 in January and Probe 5 in April was consistently largest for each norming period. This was the group of schools that was to select a total of 50 students per grade. The group of schools that was to select the fewest students per grade (43) was assigned Probe 4 in October. It was the smallest group in both October and April, but not in January. This is likely due to the incorrect selection of probes by schools within another group in January. The increase in the number of students per probe in April reflects the inclusion of Grade 1 students for the first time.

Table 3

Probe	October	Percent	January	Percent	April	Percent
1	277	16.61	273	16.46	310	16.00
2	290	17.38	250	15.07	319	16.46
3	300	17.98	286	17.24	326	16.82
4	258	15.47	327	19.71	336	17.34
5	268	16.07	257	15.49	346	17.85
6	275	16.49	266	16.03	301	15.53
Total	1668	100.00	1659	100.00	1938	100.00

Number of Students by Probe

Problems in the Data

One school (School X) forgot to administer the April probes. They were not administered until June and this delayed the final compilation of the data file. A more important issue than the delay was the decision about whether or not to include School X's June data in the April data with all other schools. The students in School X had received two more months' instruction than all other students involved in the project. In order to make my decision, I compared the April means by grade for all schools to the April means for all schools except School X (ABSX). As Table 4 shows, in all grades except Grade 5 the means with School X data included were slightly higher than without the data. The mean scores with data from School X included ranged from .02 to .97 higher than the means without the data from School X. Based on the analysis of the data summarized in Table 4, I decided to include data from School X in the development of the norms.

Table 4

Comparison of April Means

Grade	All	All But School X	Difference All-ABSX
1	15.12	14.91	.21
2	31.87	31.53	.35
3	36.95	36.50	.45
4	36.66	36.02	.64
5	38.68	38.85	17
6	58.27	58.25	.02
7	58.68	57.71	.97

Main Analysis

Descriptive Statistics by Grade

A summary of the descriptive statistics by grade and by norming period for this project is provided in Table 5. For each grade at each norming period the data for all six probes used have been aggregated.

Table 5

Descriptive Statistics

		Mean	S D	Min	Max	Skew	Kurtosis
	October						
Grade 1	January						
	April	15.12	10.28	0	63	1.69	3.93
	October	13.90	10.29	0	59	1.41	2.42
Grade 2	January	25.14	13.73	1	77	0.76	0.68
	April	31.87	15.18	1	80	0.27	-0.23
	October	21.20	12.10	0	56	0.62	-0.05
Grade 3	January	30.86	13.33	0	66	0.30	-0.62
	April	36.95	14.41	0	81	0.12	-0.35
	October	19.38	10.49	1	71	0.87	1.67
Grade 4	January	31.74	15.89	1	87	0.61	0.20
	April	36.53	17.95	0	134	0.89	2.49
	October	22.23	11.54	1	82	1.16	2.42
Grade 5	January	34.45	16.48	5	98	0.68	0.80
	April	38.70	18.76	0	102	0.53	0.13
	October	43.41	19.03	5	102	0.75	0.32
Grade 6	January	55.47	21.33	13	124	0.55	0.06
	April	58.27	23.93	7	129	0.45	-0.24
	October	47.54	23.15	4	132	0.61	0.38
Grade 7	January	55.36	24.96	8	148	0.58	0.47
	April	58.68	24.97	5	159	0.53	0.75

The data are slightly positively skewed across all grades and norming periods. This is not unexpected. It is impossible for scores on probes to be less than zero; however, across grade levels at each norming period, several students scored very high. In a normal distribution the kurtosis is zero. Most of the kurtosis values in Table 5 are near zero. The several cases of larger positive kurtosis values occur with corresponding larger positive values of skew. In examining the box plots attached in Appendix F, the six highlighted skew and kurtosis values in Table 5 are the only cases where one or more students earned an extreme score. SPSS defines an extreme score as one with a value "more than 3 box lengths from the upper or lower edge of the box" (SPSS, 1999). On the box plots circles and asterisks identify outliers and extreme scores respectively, with the adjacent number denoting the record number in the SPSS data file.

Figure 1 presents the mean and standard deviation data from Table 5 in chart format. You will note in Figure 1 that the spacing between Grades 1 and 2 on the X-axis is different from all other Grades. This is because Grade 1 students were only tested once. The grade numbers represent the first norming period for each grade.



Figure1. Mean and SD by grade and norming period.

In this visual presentation it is clear that in Grades 2 through 7 there was growth in students' computational ability throughout the school year. The growth is greatest in the early years and gradually diminishes as the students reach the upper grades. However, the jagged line for mean scores indicates an interesting phenomenon that is present between most grades. The October score for all grades except Grade 6 is lower than the April score of the previous grade.

With few exceptions standard deviations also increase with grade level. Standard deviations range from 10 to 25. This increase can be explained. The number of items on each probe remains constant at 30 (except in a few individual cases where an extra probe was used), but because scoring is based on the number of correct digits, the total possible score increases with each grade level.

Reliability of Measures

Previous research done on CBM measures of reading fluency and written expression has reported that CBM measures have demonstrated stability over time and across probes. The 1995-1996 norming project for reading fluency and written expression endorsed this stability (School District #57, 1996a). However, very little research has yet been done on CBM in math. The results of the Pearson correlation in Table 6 for correct digits scored compared between norming periods are stable. They are indications of both stability over time (6 months), and equivalence of the probes. As stability is present across groups, it can be assumed that results would be stable for an individual student. This is evidence that the probes are indeed measuring mathematics computational skills.

Table 6

	Pearson Correlation for <u>CDs Scored</u> Scores between Norming Periods									
Grade	r Oct-Jan	r Jan-Apr	r _{Oct-Apr}							
1	_	_	_							
2	.71	.73	.63							
3	.71	.74	.65							
4	.68	.74	.65							
5	.53	.63	.59							
6	.58	.69	.45							
7	.68	.73	.60							

Coefficients of Stability and Equivalence

The correlations shown in columns two and three of Table 6 are both for three-month intervals. The correlation shown in the fourth column is for a six-month interval. As one would expect, the correlations for the three-month intervals are higher overall than for the six-month interval. The only exception is at the Grade 5 level. In comparing the two three-month intervals (October – January and January – April), the correlations for the October - January interval are consistently slightly lower than for the January - April interval. This may reflect the overall drop in scores in October as described earlier when observing the mean scores in Figure 1.

Analysis of Probe Difficulty

The analysis of probe difficulty is of prime importance in this project. If the probes are not of similar difficulty, they cannot be used to assess student progress unless procedures are modified to take this into account. If a student were to be tested using an easier probe after a more difficult one, the measure of the progress would be exaggerated. Conversely, an underestimation of progress would occur if a more difficult probe were used after an easier one. As in the 1995-96 norming project, three techniques were used to analyze the probes for difficulty. First, the probe difficulties were examined at each grade level using a one factor ANOVA. The ANOVA was followed by a Scheffé post hoc comparison (Glass & Hopkins, 1996, pp. 458-459) in cases where the results of the ANOVA omnibus test indicated a need to compare the individual probes. This statistical test was conducted with $\alpha < .05$. The Scheffé post hoc test was selected as it provides "a relatively low number of false positives. It is not as likely to claim probes are of different difficulties when they are in fact of equal difficulty" (MacMillan, 1996). Second, where significant differences were found using the ANOVA and Scheffé post hoc test, the rank order placement of each probe was compared over norming periods within grade. Even if a statistical difference is found, changes in the rank order of the probe over the three norming periods, indicates that it is not different from the others. Third, the box plots for each grade level at each norming period were examined for a lack of overlap of the boxes. As described by Glass and Hopkins (1996), if all boxes on a box plot overlap, this is evidence that there is no difference. This is a very conservative test for evidence of probe differences.

Analysis of Probe Difficulty Using ANOVA and Rank Order

In Tables 7-13 the notation "ns" is used to indicate no statistically significant differences, while the notation "sig" is used to indicate significant differences have been found using the Scheffé post hoc comparison with $\alpha < .05$. A brief interpretation is provided after each table. <u>Table 7</u>

Grade 1 CD	Probe OCT	Probe JAN	Probe APR	Mean OCT	Mean JAN	Mean APR
			ns			ns
			2		in the second second second	12.43
			3			13.90
			5			14.41
			6			15.47
			4			15.94
			1			18.78

Math Probe	Differences	Across 1	Norming	F	eriods	- 1	Grade	1
------------	-------------	----------	---------	---	--------	-----	-------	---

No probes were judged significantly different at the Grade 1 level.

Table 8

Grade 2 CD	Probe OCT	Probe JAN	Probe APR	Mean OCT	Mean JAN	Mean APR
	sig	sig	ns	sig	sig	ns
	1	3	6	11.26	18.94	28.42
	4	2	3	12.00	22.74	29.00
	2	5	4	12.73	24.23	30.38
	3	1	2	13.22	24.80	31.73
	6	4	5	13.76	29.58	33.14
	5	6	1	20.77	29.77	38.77

Math Probe Differences Across Norming Periods - Grade 2

At the Grade 2 level, based on the ANOVA and the Scheffé post hoc tests, during the October and January norming periods there is a significant difference between the probes. Probe 1 ranges from being the most difficult in October to the easiest in April. However, based on a thorough comparison of rank order placement of the probes over the three norming periods, no probes were judged significantly different at the Grade 2 level.

Table 9

Grade 3 CD	Probe OCT	Probe JAN	Probe APR	Mean OCT	Mean JAN	Mean APR
	sig	sig	sig	sig	sig	sig
	4	5	6	14.51	26.95	31.05
	1	2	2	15.48	27.09	34.42
	6	1	4	19.78	27.24	36.73
	5	4	5	23.09	28.51	38.38
	2	6	3	26.18	36.36	39.51
	3	3	1	27.14	39.00	41.11

Math Probe Differences Across Norming Periods - Grade 3

At the Grade 3 level, based on the ANOVA and the Scheffé post hoc tests, there is a significant difference in the probes for all three norming periods. In comparing rank order, Probe 2 was one of the easiest probes in October and by January and April was one of the most difficult. Probe 3 was the easiest in October and January and the second easiest in April.

Table 10

Grade 4 CD	Probe OCT	Probe JAN	Probe APR	Mean OCT	Mean JAN	Mean APR
	ns	sig	ns	ns	sig	ns
	2	5	3	18.33	25.86	31.42
	6	3	6	18.46	27.94	34.67
	5	2	5	18.51	30.12	35.72
	3	6	2	19.38	34.07	36.85
	4	4	4	20.32	34.38	37.72
	1	1	1	21.40	37.20	42.89

Math Probe Differences Across Norming Periods - Grade 4

At the Grade 4 level, based on the ANOVA and the Scheffé post hoc tests, there is only a significant difference in the probes during the January norming period. In comparing rank order, Probes 1 and 4 are consistently significantly easier. All other probes change rank order at least once during the three norming periods.

Table 11

Grade 5 CD	Probe OCT	Probe JAN	Probe APR	Mean OCT	Mean JAN	Mean APR
	sig	sig	sig	sig	sig	sig
	5	5	5	15.62	27.58	31.16
	1	3	3	20.69	31.50	37.44
	3	4	.6	21.59	34.60	37.65
	2	2	4	23.53	36.07	39.29
	6	6	1	25.06	36.73	41.20
	4	1	2	26.86	39.92	45.65

Math Probe Differences Across Norming Periods - Grade 5

At the Grade 5 level, based on the ANOVA and the Scheffé post hoc tests, there is a significant difference in the probes for each norming period. However, in comparing rank order, Probe 5 is the only probe that maintains the same order across norming periods. All other probes change rank order at least once during the three norming periods.

Table 12

Grade 6 CD	Probe OCT	Probe JAN	Probe APR	Mean OCT	Mean JAN	Mean APR
	sig	sig	sig	sig	sig	sig
	5	5	5	35.84	42.74	48.51
	4	6	6	37.33	50.96	49.09
	6	4	4	40.13	52.98	52.27
	1	3	3	45.43	58.02	60.22
	3	2	1	49.82	61.68	65.57
	2	1	2	50.50	66.82	74.57

Math Probe Differences Across Norming Periods - Grade 6

At the Grade 6 level, based on the ANOVA and the Scheffé post hoc tests, there is a significant difference in the probes for all three norming periods. Probes 4, 5 and 6 are consistently more difficult than Probes 1, 2 and 3. This is the only grade level with such a well-defined split in the rank order.

Table 13

Grade 7 CD	Probe OCT	Probe JAN	Probe APR	Mean OCT	Mean JAN	Mean APR
	sig	sig	ns	sig	sig	ns
	4	2	2	39.14	44.44	52.95
	2	5	6	40.46	48.37	54.23
	1	3	1	45.37	56.14	57.29
	3	1	3	48.80	57.64	57.33
	6	6	4	53.95	60.30	63.43
	5	4	5	57.91	62.60	65.41

Math Probe Differences Across Norming Periods - Grade 7

At the Grade 7 level, based on the ANOVA and Scheffé post hoc tests, the probes are significantly different at the October and January norming periods. However when comparing rank order of the probes no consistent pattern is found. Probe 5 is the easiest in October and April, but one of the most difficult in January. No probes were judged significantly different at the Grade 7 level.

Analysis of Probe Difficulty Using Box Plots

The final test for differences in probe difficulty was the analyses of the box plots for a lack of overlap of the boxes. In order to test the box plots a ruler is placed straight across all boxes. If a straight line can be drawn that falls within all boxes, there is judged to be no difference. If one or more of the boxes falls outside the line, there is a difference.

An example of a box plot is shown in Figure 2; the remaining box plots are attached in Appendix F. The January Grade 4 box plot is a good example of a box plot that does not substantiate the differences indicated by the ANOVA and Scheffé post hoc tests. A straight line can easily be drawn through all six boxes. This indicates that the probes for Grade 4 during the January norming period can be considered equal. A slight lack of overlap was found on the box plot for the Grade 6 April probes (see Appendix F).



Figure 2. January Grade 4 box plot.

Summary of Probe Difficulty Differences

As indicated in the interpretations of the tables, at the Grade 4, 5, and 6 levels there was evidence of probe differences using an ANOVA and the Scheffé post hoc tests, and comparing rank order of the probes in the tables. However, in considering the three techniques for comparing probe differences, all probes can be considered equal. The lack of overlap on the April Grade 6 probes is not of significant concern as no difference was found on the same probes on the October and January box plots.

Creation of the Norming Tables

Smoothing

For each grade level across norming periods, tables were created in which raw scores were converted to percentiles. These are included in Appendix G. The percentile data were then converted into chart format. From the chart format, a manual smoothing process that eliminated minor inconsistencies in the data was performed. This was done by removing any overlap in the tails at either end of the charts and by maintaining approximate uniform spacing between any two lines on a chart. An example of a chart showing smoothed data is included in Figure 3. The remaining charts are attached in Appendix H. The gaps between the lines indicate growth and are seen at all grade levels between norming periods. Growth is generally greater between Fall and Winter than between Winter and Spring. Also the amount of growth between norming periods is greater for the younger grades. This is indicated by larger spaces between lines on the charts for the lower grades. The greatest amount of smoothing was required in Grades 6 and 7 where at several percentile levels the Winter scores were slightly higher by a mark or two than Spring scores.



Figure 3. Smoothed percentile data for Grade 7.

Percentile data on the smoothed charts was converted into table format to create the norming tables for SD57 use. The designations of below average, average, and above average were determined by using scores that fell below the 25th percentile, between the 25th and 75th percentile, and above the 75th percentile respectively. Table 14 represents the Grade 7 norms that correspond to the data in Figure 3. The norming tables for the remaining grades are attached in Appendix I.

Table 14

CBM Grade 7 norms.

	C	RADE SEVEN Co	orrect Digits Sco	ored
	Fall	Winter	Spring	
Percentile	CD	CD	CD	Description
99	123	130	134	
95	86	99	110	
90	78	86	90	Well Above Average
85	74	80	84	
80	69	75	78	
75	63	69	72	Above Average
70	59	66	70	
65	54	64	67	
60	51	60	65	
55	48	57	60	
50	45	53	57	Average
45	42	50	53	
40	38	47	51	
35	36	44	48	
30	32	41	46	
25	30	37	41	Below Average
20	27	33	38	
15	24	29	32	
10	20	24	27	Well Below Average
5	14	17	20	
1	5	7	9	

DISCUSSION

Within the scope of this project three variables were subjected to statistical analysis. They were the score measured by number of correct digits, grade level, and norming period. The goal of the project was to establish local norms for computational math for SD57. However, in completing the statistical analysis required to establish these norms, reliability of the measures and probe difficulty were also analyzed. The data also provided information on students' growth in learning that confirmed the pattern established in SD57's earlier project to establish local norms for reading fluency and written expression.

Issues Raised by Data

Students' Growth and Summer Effect

As was found in the 1995-96 project to develop local norms for reading fluency and written expression, students improved in their ability to compute mathematically over the course of the school year. This improvement within each grade level (except Grade 1 where there was only one norming period) is evident by comparing the mean scores in Table 5, by the growth line in Figure 1, and by the spaces between the lines on the charts in Appendix H. This growth is greatest in the early grades and gradually diminishes at the upper grades. This is most likely because the four computational operations (addition, subtraction, multiplication and division) are new concepts for students in the early grades and they are learning them rapidly. At the upper grade levels most students have already reached their own level of proficiency in performing the operations. At this level the computations become more complex as fractions and decimals are introduced. Another reason for relatively little growth shown by Grade 6 and 7 students between January and April could be due to some students shutting down and not trying their best at that point in the school year.

However, in examining students' growth in computation ability, it is interesting to note that in all but one case, the October mean score for each grade was lower than the previous grade's April mean score. This was most clearly evident in the jagged growth line in Figure 1. The reason for this drop in scores at the beginning of most grades can be explained in part by the increasing difficulty of the probes. A Grade 4 student writing a Grade 4 probe in October will be encountering items that test the entire Grade 4 curriculum. Many of the concepts included on the probe will not have been taught this early in the school year. Interestingly, the same jagged-line effect was found for both the reading fluency and written expression scores in the 1995-96 project (School District #57, 1996a). However only the reading fluency probes increased in difficulty with grade level; the written expression probes did not. The probes used for written expression were identical at each grade level and the total number of words written was used to measure the amount of writing. The jagged-line effect was still present for written expression as the amount of writing dropped each October as compared to April of the previous grade. In this case it cannot be attributed to increasing difficulty of the probes. In the reading and writing norming project the drop in scores each October was attributed to a summer effect (School District #57, 1996a). This explanation takes into account the fact that students lose ground over the summer months when they are not practicing the skills they have learned in school. In the case of both reading fluency and math computation the lower October skills may be partially due to a summer effect, but the increase in the difficulty of the material must also be considered. Reliability of the Measures

The results of the Pearson correlations presented in Table 6 provide evidence that the probes are stable over both time and testers. Six different probes at different grade levels over three norming periods were administered as much as 6 months apart to groups of approximately

275 students from throughout SD57. Additionally, these probes were scored by at least 50 different raters. The Pearson correlations for this project are comparable to those from the 1995-96 reading fluency and written expression project. For the six-month period from October to April, the median coefficient of stability for reading fluency was .83 with a range from .77 to .86. For written expression the median was .59 with a range of .48 to .62 (School District #57, 1996a). In this project for math computation the median was .63 with a range from .45 to .65. As expected, slightly higher results were found in the two three-month periods October to January and January to April in all three curricular areas. The results for math are closer to those of written expression than to reading fluency. A possible explanation for this is that both math and writing are skills that need to be practiced in order to be maintained. Although this may also be true for reading, I believe it is not so to the same extent. Once children learn to read they do not often forget how. Also, over the summer children are much more likely to continue to read than to do math or write.

As I indicated in the introduction, both Deno (1992) and Marston (1989) believe that not enough research has been conducted on the reliability of CBM in the area of mathematics. This project adds to this identified lack of research in this area. Also, when the results of this project are combined with the results from the 1995-96 project to establish local norms in reading fluency and written expression, additional evidence on the reliability of CBM measures in general is provided.

Probe Difficulty

Unless it could be determined that the six probes for each grade level were of equal difficulty, there would be no point in creating norming tables for the probes. Three techniques were used to compare the probes within each grade level: a one factor ANOVA followed by the
Scheffé post hoc test, comparison of rank order over the three norming periods, and examinations of the box plots for a lack of overlap. Probes were judged to be different only after all three techniques had been applied. At the Grade 3, 4, 5, and 6 levels differences in some probes were found after applying the one factor ANOVA followed by the Scheffé post hoc test and comparing the rank order of the probes. An examination of the box plots for these identified grade levels showed a slight lack of overlap only for the Grade 6 April norming period. The box plots for the Grade 5 October norming period slightly overlapped. The Grade 5 and 6 probes were judged to be equal because there was overlap on the other two box plots for each grade. However, in my technical report to the school district, I did recommend that they have members of the math committee look closely at the probes at the Grade 5 and 6 levels to see if the committee members noticed any differences in the probes (Walraven & MacMillan, 2000).

When using CBM to assess individual students, a common practice is to administer three probes within a short period of time (one to three days) and compare the median score to local norms (School District #57, 1996a; Shinn, 1989). This practice reduces the effect of any slight differences in probes.

Concluding Statement

The purpose of this project was to develop local CBM math norms for SD57. The probes are not comprehensive math tests. They are strictly measures of computational skills in a timed setting. However, they are still useful. Computational ability is a cornerstone of overall mathematics ability. Students who can compute well are more likely to be better math students than those who cannot. The probes are quick and inexpensive to administer. An entire class can be tested on a probe in less than ten minutes. They can be used to identify students who may require more practice in computation, or who may require further testing. They definitely meet the school district's requirement for simple and informal measures to monitor and evaluate student progress (School District #57, 1996a).

References

Brigance, A. H., & Hargis, C. H. (1993). <u>Educational assessment: Insuring that all</u> students succeed in school. Springfield IL: Thomas Books.

Deno, S. L., (1985). Curriculum-based measurement: The emerging alternative. Exceptional Children, 52, 219-232.

Deno, S. L. (1992). The nature and development of curriculum-based measurement. Preventing School Failure, 36 (2), 5-10.

Deno, S. L. & Fuchs, L. (1988). Developing curriculum-based measurement systems for data-based special education problem solving. In E. L. Meyer, G. A. Vergason, & R. J. Whelan (Eds.), <u>Effective instructional strategies for exceptional children</u> (pp. 481-504). Denver, CO: Love.

Fewster, S. A. (2000). <u>School-based evidence for the validity of curriculum-based</u> <u>measurement norms in School District #57</u>. Unpublished master's thesis, University of Northern British Columbia, Prince George, British Columbia, Canada.

Filemaker Pro 4.1 for Windows and Macintosh [Computer Software]. (1998). Santa Clara, CA: Filemaker.

Fuchs, D. & Fuchs, L.S. (1992). Identifying a measure for monitoring student reading progress. <u>School Psychology Review</u>, 21, 45-58.

Glass, G. V. & Hopkins, K. D. (1996). <u>Statistical methods in education and psychology</u>. (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Hedekar, L. (1997). <u>The effects of month of birth and gender on elementary reading and</u> writing fluency scores using curriculum based measurement. Unpublished master's thesis, University of Northern British Columbia, Prince George, British Columbia, Canada.

Howell, K. W., Fox, S. L., & Morehead, M. K. (1993). <u>Curriculum-based evaluation:</u> <u>Teaching and decision making</u> (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Kaminski, R. A., & Good, R. H., III. (1998). Assessing early literacy skills in a problem solving model: Dynamic indicators of basic early literacy skills. In M. R. Shinn (Ed.), <u>Advanced applications of curriculum-based measurement (pp. 113-142)</u>. New York: Guilford Press.

Knutson, N., & Shinn, M. R. (1991). Curriculum-based measurement: Conceptual underpinnings and integration into problem-solving assessment. Journal of School Psychology, 29, 371-393.

MacMillan, P. (1995). [Interim technical report - October norming results]. Unpublished report.

MacMillan, P. (1996). [Curriculum based measurement norming study – final report]. Unpublished report.

MacMillan, P. (2000). Simultaneous measurement of reading growth, gender, and relative age effects: Many faceted Rasch applied to CBM reading scores. Journal of Applied Measurement, 1, 393-408.

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), <u>Curriculum-based measurement:</u> <u>Assessing special children</u> (pp. 18-78). New York: Guilford Press.

Moore, A. D., & Young, S. (1997). Clarifying the blurred image: Estimating the interrater reliability of performance assessments. (ERIC Document Reproduction Service No. ED414 319)

Salvia, J. & Ysseldyke, J. E. (1991). Assessment (5th ed.). Boston: Houghton Mifflin.

School District #57. (1996a). <u>Guidebook for the use of curriculum based measurement in</u> <u>School District #57</u>. Prince George, BC: School District #57.

School District #57. (1996b). <u>School support services: Practices, organization, principles</u>. Prince George, BC: School District #57.

School District #57. (1999). <u>CBM math (calculation) norming project, 1999-2000.</u> Prince George, BC: School District #57.

Shinn, M. R. (1988). Development of curriculum-based local norms for use in special education. <u>School Psychology Review</u>, 17, 61-80.

Shinn, M. R. (1989). Identifying and defining academic problems: CBM screening and eligibility procedures. In M. R. Shinn (Ed.), <u>Curriculant-based measurement: Assessing special children</u> (pp. 90-129). New York: Guilford Press.

Shinn, M. R., & Bamonto, S. (1998). Advanced applications of curriculum-based measurement: "Big Ideas" and avoiding confusion. In M. R. Shinn (Ed.), <u>Advanced applications</u> of curriculum-based measurement (pp. 1-31). New York: Guilford Press.

Shinn, M. R., Nolet, V., & Knutson, N. (1990). Best practices in curriculum-based measurement. In A. Thomas & J. Grimes (Eds.) <u>Best practices in school psychology-II</u> (pp. 287-307). Washington, DC: National Association of School Psychologists.

SPPS Graduate Pack 9.0 for Windows [Computer software]. (1999). Chicago: SPSS.

Tindal, G., Germann, G., & Deno, S. L. (1983). <u>Descriptive research on the Pine County</u> <u>norms: A compilation of findings</u> (Research Report No. 132). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.

Tindal, G., Marston, D., & Deno, S. L. (1983). <u>The reliability of direct and repeated</u> <u>measurement</u> (Research Report No. 109). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.

Walraven, G. B., & MacMillan, P. (2000). [Technical report of the curriculum based measurement (Math) norming project]. Unpublished report.

Worthen, B. R., Borg, W. R., & White, K. R. (1993). <u>Measurement and evaluation in the</u> schools. New York: Longman.

Appendix A

Problem Solving Approach and Process Chart

CHART 1

PROBLEM SOLVING APPROACH AND PROCESS



 Adapted from Heartland Education Agency 11 Division of Special Education - Iowa Appendix B

•

Grade 7 CBM Math Probe

				Grade 7, probe 1	
PEN	_ Name	· ·		<u>CD</u>	
15637 <u>- 9859</u> 5778	$(+9) - (-4) = \frac{+13}{13}$ or 13	19 152 152 1710 1862	36 47 <u>+ 12</u> 95	22 72)1651 144 211 144 67	
(4)	(3)	(11)	(2)	(13)	3
4.2 + 24.7 = <u>38</u> 9	12.2 .6)7.32 6 13 12 13 12 13	15 26)403 26 143 130 13	(-7) - (-2) = <u>-5</u>	10.4 + 9.12 = <u>19.52</u>	
(4)	(14)	(12)	(2)	(5)	37
.87 4)3.48 32 28 28 0	57 ÷ <u>19</u> = 3	(+7) x (-3) = <u>-31</u>	$\frac{24}{6} = 4$	$(+5) \div (-1) = -5$	
(10)	(2)	(3)	(1)	(2)	8

587 <u>x 37</u> 4109 17610 21719	4.2 - 1.58 = <u>3.62</u>	8637 <u>- 2918</u> 57/9	12% x 50 = 6	179.4 <u>+ 25.6</u> 205.0 or 205	
(14)	(4)	(4)	(1)	(5)	โล
8437 <u>+5976</u> 14413	(+8) + (-2) = <u>+6</u>	<u>70</u> x 70 = 4900	(+3) + (+4) = <u>77</u> or 7	19 32)608 32 288 288 288 288	مراجعه والمحافظ
(5)	(2)	(2)	(2)	(11)	22
2303 <u>- 1805</u> 498	.174 = <u>17.4</u> %	14 x <u>50</u> = 700	3600 ÷ 60 = <u>60</u>	20% of 70 = <u>14</u>	
(3)	(4)	(2)	(2)	(2)	13

.

Appendix C

Student Selection/ Probe Sequence Table

Student Selection/Probe Sequence

Elementary Schools	(per grade)	Probe #	Elementary Schools	(per grade)	Probe #
Morfee	8	1	K. G. V. English	3	3
Austin Road	9	1	Haldi Road	2	3
Immersion	2	1	McLeod Lake	1	3
Vanway	6	1	Bear Lake	1	4
Springwood	7	1	Blackburn	11	4
Van Bien	5	1	Quinson	6	4
Meadow	4	1	Ron Brent	5	4
Seymour	3	1	North Nechako	6	4
Montessori	Gr. 1-4, Gr. 2-1	1	Mountain View	5	4
Red Rock	1	1	Harwin	4	4
Dunster	1	1	Central Fort George	3	4
Westwood	10	2	Giscome	2	4
Carney Hill	8	2	Valemount	6	5
Beaverly	7	2	Hart Highlands	9	5
Lakewood	6	2	Heritage	8	5
Glenview	5	2	Peden Hill	7	5
Ft. George South	3	2	Gladstone	6	5
Edgewood	2	2	McBride Centennial	5	5
Dome Creek	0	2	Buckhorn	4	5
Pineview	6	2	Malaspina	8	6
Hixon	1	2	Foothills	7	6
Highglen	3	3	Hart Highway	7	6
Montessori	5	3	Spraceland	6	б
Mackenzie	7	3	Immersion	4	б
Highland	6	3	Nukko Lake	5	6
College Heights	5	3	Pinewood	4	6
Immersion	4	3	Shady Valley	3	6
Southridge	9	3	Salmon Valley	2	6
Wildwood	5	3			
				-	

Appendix D

Random Selection of Students Table

A

Random Selection of Students

Grade	Students to be used in Norming Sample
Grade 1	10, 11, 15, 18, 19, 23, 27, 28, 30, 34, 41
Grade 2	4, 5, 7, 8, 16, 18, 20, 22, 23, 35, 47
Grade 3	2, 3, 16, 22, 26, 27, 33, 37, 41, 45, 51
Grade 4	4, 11, 12, 13, 25, 26, 30, 39, 40, 43, 50
Grade 5	4, 5, 7, 18, 20, 21, 23, 28, 34, 36, 42
Grade 6	7, 8, 12, 13, 15, 19, 23, 28, 30, 39, 44
Grade 7	5, 8, 9, 13, 21, 26, 37, 39, 41, 45, 53

Adapted from School District No. 57 CBM Math (Calculation) Norming Project, 1999 - 2000 Appendix E

Letters of Permission

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

UNBC

3333 University Way, Prince George, BC V2N 4Z9

Dr. Alex Michalos Chair, UNBC Ethics Review Committee Tel: (250) 960-6697 or 960-5820 Fax: (250) 960-5746 E-mail: michalos@unbc.ca

UNBC Ethics Committee

July 4, 2000

Gail Walraven 2331 McBride Crescent Prince George, BC V2M 1Z8

Proposal: EP20000615.48

Dear Ms. Walraven:

Thank you for submitting your proposal entitled, "CBM Math Norming Study".

Your proposal has been approved and you may proceed with your research. Please ensure however, that you receive written permission from School District No. 57 to use the data as part of your thesis/project.

If you have any questions, please feel free to contact me.

Sincerely,

Alex Michalos ⁽ Chair, UNBC Ethics Review Committee



SCHOOL DISTRICT NO. 57 (PRINCE GEORGE)

1894 Ninth Avenue, Prince George, B.C. V2M 1L7

Phone: (250) 561-6800 Fax: (250) 561-6801 www.schdist57.bc.ca

January 12, 2001

Gail Walraven Vice-Principal Kelly Road Zone C/O Austin Road Elementary

Dear Gail:

This letter is to confirm that after having discussed your research proposal with you, District permission for you to access Curriculum Based Measurement (C.B.M.) data is hereby granted. We wish you well with your U.N.B.C. "C.B.M. Math Norming Study."

Sincerely,

Norm Monroe Director School Services

NM/hg

Appendix F

Box Plots for all Norming Periods



OCTPROBE



OCTPROBE















JANPROBE



JANPROBE



JANPROBE



JANPROBE



JANPROBE



GRADE= 1



APRPROBE



APRPROBE



APRPROBE









APRPROBE

Appendix G

Percentile Scores – Raw Data Grades 1-7

		GRADE ONE Cor	rect Digits Scored	1
	Fall	Winter	Spring	
Percentile	CD	CD	CD	Description
99			56	
95			35	
90			28	Well Above Average
85			24	
80			20	
75			19	Above Average
70			18	
65			17	
60			15	
55			14	
50			13	Average
45			12	
40		-	11	
35			10	
30			9	
25			8	Below Average
20			7	
15			6	
10			5	Well Below Average
5			4	
1			1	
n =			282	Number in Sample

N.B. Grade One students were tested only once, during the April norming period.

		GRADE Two Cor	rect Digits Scored	1
	Fall	Winter	Spring	
Percentile	CD	CD	CD	Description
99	50	69	74	
95	36	50	55	
90	27	43	51	Well Above Average
85	23	40	48	
80	22	37	46	
75	19	33	44	Above Average
70	17	31	41	
65	15	29	38	
60	14	27	35	
55	12	25	34	
50	11	23	31	Average
45	10	21	29	
40	9	19	27	
35	8	18	26	
30	7	17	23	
25	6	15	20	Below Average
20	6	13	17	
15	5	11	15	
10	4	9	12	Well Below Average
5	2	6	8	
1	0	2	3	
n =	278	278	275	Number in Sample

		GRADE THREE Co	orrect Digits Scor	ed
	Fall	Winter	Spring	
Percentile	CD	CD	CD	Description
99	53	61	71	
95	45	54	61	
90	40	51	56	Well Above Average
85	35	46	53	
80	30	43	50	
75	28	40	47	Above Average
70	26	38	45	
65	24	36	43	
60	22	34	41	
55	20	32	39	
50	19	28	37	Average
45	18	27	35	
40	17	25	32	
35	16	24	30	
30	14	22	28	
25	13	21	26	Below Average
20	11	19	24	
15	9	17	22	
10	6	16	19	Well Below Average
5	3	11	15	
1	0	4	4	
n =	280	277	279	Number in Sample

	Fall	Winter	Samina	
D	rall	winter	Spring	
Percentile	CD	CD	CD	Description
99	51	75	84	
95	39	60	69	
90	33	54	59	Well Above Average
85	29	48	55	
80	27	46	51	
75	26	41	47	Above Average
70	25	39	44	
65	22	36	42	
60	21	34	38	
55	20	32	36	
50	19	29	33	Average
45	17	28	32	
40	16	26	31	
35	14	24	29	
30	13	23	27	
25	12	21	25	Below Average
20	10	18	22	
15	8	16	18	
10	6	12	14	Well Below Average
5	5	9	10	
1	3	2	3	
n =	278	276	274	Number in Sample

		GRADE FIVE Cor	rect Digits Score	d
	Fall	Winter	Spring	
Percentile	CD	CD	CD	Description
99	60	88	91	
95	45	63	72	
90	40	55	63	Well Above Average
85	34	50	59	
80	31	48	54	
75	28	46	50	Above Average
70	26	43	47	
65	24	40	45	
60	23	37	43	
55	21	35	39	
50	20	32	37	Average
45	19	30	35	
40	17	29	33	
35	16	26	30	
30	15	25	27	
25	14	22	24	Below Average
20	13	20	21	
15	12	18	19	
10	9	13	15	Well Below Average
5	8	11	12	
1	3	6	5	
n =	281	276	277	Number in Sample

		GRADE SIX Com	rect Digits Scored	
	Fall	Winter	Spring	
Percentile	CD	CD	CD	Description
99	97	116	123	
95	84	96	99	
90	71	84	91	Well Above Average
85	62	77	86	
80	58	72	80	
75	54	68	74	Above Average
70	50	65	69	
65	48	63	65	
60	45	60	61	
55	43	58	58	
50	40	54	56	Average
45	38	50	53	
40	36	47	50	
35	34	45	46	
30	32	42	44	
25	29	39	40	Below Average
20	27	36	37	
15	25	34	34	
10	23	29	29	Well Below Average
5	17	24	23	
1	9	19	15	
n =	276	277	276	Number in Sample

T	Eall	Winter	Carries	
	Fall	winter	Spring	
Percentile	CD	CD	CD	Description
99	123	130	134	
95	86	99	102	
90	78	88	89	Well Above Average
85	74	81	82	
80	69	75	78	
75	63	70	72	Above Average
70	59	66	70	
65	54	64	67	
60	51	60	65	
55	48	57	60	
50	45	53	57	Average
45	42	50	53	
40	38	47	51	
35	36	44	48	
30	32	41	46	
25	30	37	41	Below Average
20	27	33	38	
15	24	29	32	
10	20	25	26	Well Below Average
5	14	20	19	
1	6	9	9	
n =	275	275	275	Number in Sample

Appendix H

Charts of Smoothed Percentile Data Grades 1-6












Appendix I

CBM Math Norms Grades 1-6

		GRADE ONE Con	rect Digits Scor	red
	Fall	Winter	Spring	
Percentile	CD	CD	CD	Description
99			56	
95			35	
90			28	Well Above Average
85			24	
80			20	
75			19	Above Average
70			18	
65			17	
60			15	
55			14	
50			13	Average
45			12	
40			11	
35			10	
30			9	
25			8	Below Average
20			7	
15			6	
10			5	Well Below Average
5			4	
1			1	

N.B. Grade One students were tested only once, during the April norming period.

		GRADE Two Cor	rect Digits Score	ed
	Fall	Winter	Spring	
Percentile	CD	CD	CD	Description
99	50	65	74	
95	36	50	60	
90	27	43	51	Well Above Average
85	23	40	48	
80	21	37	46	
75	19	33	44	Above Average
70	17	31	41	
65	15	29	38	
60	14	27	36	
55	12	25	34	
50	11	23	31	Average
45	10	21	29	
40	9	19	27	
35	8	18	26	
30	7	17	23	
25	6	15	20	Below Average
20	5	13	17	
15	4	11	15	
10	3	9	12	Well Below Average
5	2	6	8	
1	0	2	4	

	0	GRADE THREE CO	orrect Digits Sco	ored
	Fall	Winter	Spring	
Percentile	CD	CD	CD	Description
99	53	61	71	
95	45	54	61	
90	40	50	56	Well Above Average
85	35	46	53	
80	30	43	50	
75	28	40	47	Above Average
70	26	38	45	
65	24	36	43	
60	22	34	41	
55	20	32	39	
50	19	30	37	Average
45	18	27	35	
40	17	25	32	
35	16	24	30	
30	14	22	28	
25	13	21	26	Below Average
20	11	19	24	
15	9	17	22	
10	6.	14	19	Well Below Average
5	3	10	15	
1	0	2	4	

	(GRADE FOUR Co	rrect Digits Sco	red
	Fall	Winter	Spring	
Percentile	CD	CD	CD	Description
99	51	75	84	
95	39	60	69	
90	33	54	59	Well Above Average
85	29	48	55	
80	27	45	51	
- 75	26	41	47	Above Average
70	24	39	44	
65	22	36	41	
60	21	34	38	
55	20	32	36	
50	19	30	34	Average
45	17	28	32	
40	16	26	31	
35	14	24	29	
30	13	23	27	
25	12	21	25	Below Average
20	10	18	22	
15	8	15	18	
10	6	11	15	Well Below Average
5	5	8	11	
1	1	2	4	

		GRADE FIVE Con	rect Digits Scor	red
	Fall	Winter	Spring	
Percentile	CD	CD	CD	Description
99	60	80	91	
95	45	63	72	
90	40	55	63	Well Above Average
85	34	50	59	
80	31	48	54	
75	28	46	50	Above Average
70	26	43	47	
65	24	40	45	
60	23	37	43	
55	21	35	39	
50	20	32	37	Average
45	19	30	35	
40	17	29	33	
35	16	26	30	
30	15	24	28	
25	14	22	25	Below Average
20	13	19	22	
15	12	16	19	
10	9	13	16	Well Below Average
5	8	10	12	
1	3	4	5	

		GRADE SIX Con	ect Digits Score	ed
	Fall	Winter	Spring	
Percentile	CD	CD	CD	Description
99	97	116	123	
95	84	96	101	
90	71	84	91	Well Above Average
85	62	77	86	
80	58	72	80	
75	54	68	74	Above Average
70	50	65	69	
65	48	63	65	
60	45	60	63	
55	43	57	59	
50	40	53	56	Average
45	38	50	53	
40	36	47	50	
35	34	44	47	
30	32	41	44	
25	29	38	41	Below Average
20	27	34	37	
15	25	30	34	
10	23	26	29	Well Below Average
5	17	22	24	
1	9	15	18	